DOCUMENT RESUME

ED 235 477 CS 207 464

AUTHOR Black, Mary C.
TITLE Readability as an Interactive Process: An Analysis of the Syntactic and Semantic Effects on Text of Cloze Errors Associated with Variation in the Structure of Text and in Readers' Knowledge of Topic.
PUB DATE [79]
NOTE 32p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cloze Procedure; *Cognitive Processes; Error Analysis (Language); Models; *Readability; *Readability Formulas; *Reading Processes; *Reading Research; Secondary Education; Secondary School Students; *Sentence Structure
IDENTIFIERS *Reader Text Relationship

ABSTRACT

A study examined whether (1) there are untested indices of difficulty that can be used to enhance the accuracy of readability formulas, (2) the same readability formula can be used to assess the readability of text for mature and immature readers and for good and poor readers, and (3) some one model of the reading process or some synthesis of models presents a better account than others of findings on readability and perception. Subjects, 239 secondary school students, completed questionnaires on which they indicated the extent of their knowledge of and interest in each of 10 topics. Test passages on these topics were first scored to determine the total number of cloze errors; then, a sample was scored to devise a system that could be used to describe the effects of incorrect insertions on the structure and meaning of the text sentence. Results showed that sentence unit integrity tended to be maintained or created at all reading and writing levels. When the topic was familiar and the text became increasingly complex, readers who were having difficulty made an increasing percentage of errors which maintained unit structure and created a sensible alternative reading of the otherwise undeleted unit, but when text was less familiar this did not occur. Results also revealed that the processing of units occurring at the beginning of sentences resulted in the perception of a unit more frequently than the processing of units that occurred later in the sentence. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235477

Readability as an Interactive Process: An Analysis of the Syntactic and Semantic Effects on Text of Cloze Errors Associated with Variation in the Structure of Text and in Readers' Knowledge of Topic

Mary C. Black
20 W. Carey Street
Plains, PA 18705
(Phone (717) 822-5469)
Position: Activities Director,
Academic Support Center
Wilkes College
Wilkes-Barre, PA 18766
(Phone (717) 824-4651)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary C. Black

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 207 464

Justification

Readability, the relative comprehensibility of texts, has been the subject of discussion and empirical research for over sixty years. In 1935, Gray and Leary defined readability as the product of an interaction between text characteristics and reader characteristics. Although this definition has not been challenged, most readability investigations, including that by Gray and Leary, have been undertaken to develop readability formulas through the identification of those text characteristics that correlate best with a hierarchy of arbitrarily designated levels of understanding.

Both the accuracy of readability formulas and the means used to determine their accuracy have varied from one investigation to another. Of those formulas whose accuracy of prediction was determined through a product-moment correlation with a measure of comprehensibility other than another formula, only one (Vogel & Washburne, 1928) correctly predicted the comprehensibility of more than 50% of the texts used in the study.

Readability formulas are used by publishers to rank the difficulty of texts used in elementary and secondary schools, and the formulas are sometimes used in selecting college texts. In school settings, particular texts are often chosen to be read by particular students because the readability levels of the texts, determined through the application of readability formulas, are the same as the reading-levels of the students, determined through testing. It seems difficult to place high confidence in this practice because the accuracy of readability formulas appears to be seriously limited and because it is probable that the readability of a given text for a specific reader may be a function of such within-reader characteristics as knowledge of content (Chall, 1947; Robinson, 1947), and interest (Bernstein, 1955; Estes & Vaughn, 1973) as well as of the structural aspects of text indexed in readability formulas.

The purpose of readability research has been the same as the purpose of much of the research in perception, the identification of variables which influence the comprehension of text. Therefore, findings from research in readability are relevant to the determination of the perceptual reading process and, by implication, to the building of models of the reading process. For this reason, both the background and the findings of the present study were referenced to readability research, perceptual research, and model building.

There were two general purposes for this research. The first was to investigate the relative influence of the structure of text and of one within-reader characteristic, subjectively evaluated knowledge of text topic, on the comprehensibility of text, measured by the number of cloze errors on test passages. The second was to determine how variation in the structure of text, in readers' knowledge of topic, and in the extent of the readers' understanding of the test passages influenced the perception of text, inferred through analyses of the effects of cloze errors on the structure and meaning of the sentences in which they occurred. The findings were related to the following specific questions and corollaries.

Question I - Are there untested indices of difficulty that can be used to enhance the accuracy of readability formulas?

Corollary 1 - Can a readability formula be constructed to be more easily applied and more accurate than currently popular formulas?

Corollary 2 - Can readability formulas be constructed which are applicable to all topics, regardless of the level of knowledge of the reader?

Question II - Can the same readability formula be used to assess the readability of text for mature and immature readers and for good and poor readers?

Question III - Does some one model of the reading process or some synthesis of models present a better account than others of findings on readability and perception?

Terms

Cloze Test - A measure of comprehension in which every n^{th} word in text is deleted.

Exact Scoring - The method for scoring a cloze test in which an insertion is considered to be correct only if it is exactly the word deleted.

Written Level - The grade-level number given text that is comprehensible to normally achieving students in that grade in school but not in a lower grade. In this study, written level also refers to the factor in which average sentence length and the average number of syllables per hundred words were manipulated. The designation of particular written levels (fourth, sixth, eighth, and tenth) was done in accordance with ranges of these criteria in the Fry Readability Formula (1968) associated with written level as defined above.

Reading Level - A level of comprehension measured by a score on a test of comprehension and related to the reader's ability to learn from text. It is expressed as the written level that is associated with specific types of reading ability: the independent level is that at which the reader can profit from text without instruction; the instructional level is that at which the reader can profit from text with instruction; the frustration level is that at which the reader is unable to profit from text with instruction (Johnson & Kress, 1965). The last category was subdivided into the frustration and below frustration levels in an attempt to enhance the discriminability of reading levels. Reading level was an assigned factor in this study. The cut-off points for designating which cloze scores fell within particular levels were established according to criteria developed by Jones and Pikulski (1974). The cut-off point for the below frustration level was arbitrarily defined.

Topic - The factor in this research which was manipulated on the basis of readers' subjective evaluation of their knowledge of topic. There were two levels of this factor: topic of greatest knowledge and topic of least knowledge.

Syntax or Grammar - The surface phrase structure of text.

Semantics - Meaning, in the general sense of the word.

Sentoid - Any surface phrase structure which can be shown to be derived from an underlying phrase structure consisting of at least a noun phrase and a verb phrase. There can be one or more sentoids in a sentence. For example, John wanted a dog, consists of one sentoid, and, John wanted to go home, consists of two. The last sentoid, to go home, is embedded within the first, John wanted. The subject of the second sentoid, John, is present in underlying structure, but it has been deleted in surface structure.

Selection Restrictions - The restrictions imposed by context on the nature of an element, an abstractly represented word that can be inserted into context. Thus, in, The rock slept, the restriction imposed on the verb by the noun, that it must be marked to occur only in the context of an animate subject noun, is violated.

Heuristics or Strategies - The processes through which readers utilize sentential constraints in extracting meaning from text. Heuristics may be based on subconscious knowledge of the redundancies of text upon which readers, also subconsciously, base generalizations regarding the nature of sentential constraints. These generalizations may enable them to form expectancies about the nature of subsequent text, given the nature of prior text. Fodor, Bever, and Garrett (1974) posit a canonical sentoid strategy. In this, readers approach text with the expectation that the first noun encountered will be the subject noun and the first verb, its predicate.

Background

Question I and Corollaries 1 and 2

All readability formulas contain a measure of vocabulary load and sentence length (Klare, 1968). Although a few have also included other indices of difficulty, these formulas have not been widely used because the added indices did not contribute substantially to prediction (Gray & Leary, 1935) and because the counting of added text attributes made the formulas difficult to apply (Washburne & Morphett, 1938).

Bormuth (1966) investigated the relative utility of traditional readability measures and measures suggested by the then contemporary formulations in theoretical linguistics. He found that the latter measures were viable indices of difficulty, and that ratios of both traditional and nontraditional linguistic variables enhanced the prediction of comprehensibility. These findings did not generate research to develop more accurate formulas. This may have been a consequence of the fact that at the time that his findings were published, little empirical research had been done to determine the "psychological reality" of aspects of the new grammatical constructs, making it difficult to provide a rationale for testing specific sets of ratios.

Some of the structural text attributes found to be perceptually salient in research influenced by the new grammar are major and minor phrase boundaries (Garrett, Fodor, & Bever, 1966; Johnson, 1965; Levin & Kaplan, 1970; Schlesinger, 1968), sentoid boundaries (Abrams & Bever, 1969; Bever, Lackner, & Kirk, 1969; Blumenthal, 1966; Foster, 1970), and traditionally defined clause boundaries (Coleman, 1965; Levin & Kaplan, 1970; Wanat, 1976). These findings provide evidence that readability prediction might be enhanced by including in readability formulas ratios which index these text aspects. Such ratios may provide a more sensitive measure of structural complexity than average sentence length,

currently used in formulas. This speculation is consonant with Bormuth's (1966) finding that the number of nouns divided by the number of structure words, such as prepositions and conjunctions, added independently to prediction when sentence length had already been included in the prediction equation. Structure words may index the number of sentoids and clauses, and the number of nouns, sentence length. In a separate analysis on traditional parts of speech, Bormuth found that the number of pronouns divided by the number of conjunctions accounted for 66.4% of passage difficulty; this is higher predictive accuracy than has been reported for most currently used readability formulas. Pronouns may index the number of sentoids, and conjunctions, the number of clauses.

Other findings provide evidence that readers are sensitive to selection restrictions (Fodor & Garrett, 1968; Schlesinger, 1968) and to the semantic structure of text (Anderson, Reynolds, Schallert, & Goetz, 1976; Bransford & Franks, 1971; Marshall & Glock, 1978; Meyer, 1975). It does not seem possible at this time to infer potential measures to index these text attributes because little is known about the nature of the perceptually salient text attributes that facilitate their processing.

Within-reader factors, such as knowledge of topic, may have substantial impact on comprehensibility. Bormuth (1966) suggested that the improved indexing of text factors would result in the indexing of within-reader characteristics. This may be theoretically possible for readers in general because it seems likely that variation in within-reader characteristics would be associated with variation in the relative perceptual salience of various aspects of text. Thus, readers with high knowledge of some topic might be more or less sensitive to particular text attributes than readers with low knowledge of that topic. However, it seems difficult to see how such sensitivity could be indexed in a readability formula without using a measure that would reflect the extent of

a particular reader's knowledge of topic.

Question II

The question as to whether it is possible to use one formula to predict readability for readers of varying maturity and competence derives primarily from the Gray and Leary (1935) finding that readability for poor readers was predicted best by lexical factors, and readability for good readers was predicted best by structural factors. Readability formulas are meant to rank texts along a continuum of difficulty defined in terms of the grade-levels in school of normally achieving, competent readers. Because the poor readers in the Gray and Leary study were those with low scores on the measure of comprehension of the passages used in the study, the finding must be cautiously interpreted. Findings from perceptual research are also equivocal in relation to this question.

Good readers and mature readers exhibit similar eye-movement behaviors as do poor and immature readers. Good and mature readers have longer eye-voice spans (Buswell, 1920; Fairbanks, 1937; Levin & Kaplan, 1970), process more words in a single fixation (Buswell, 1920; Judd, 1918; Levin & Kaplan, 1970), regress less frequently (Buswell, 19201 Fairbanks, 1937; Tinker, 1965), and make more sensible corrections or oral reading errors (Fairbanks, 1937; Weber, 1970) than poor and immature readers. These findings have been interpreted as evidence that because poor readers do not read for meaning, they process text word by word (Buswell, 1920; Anderson & Swanson, 1937). The findings can also be attributed to difficulty in decoding words. However, they do not provide proof that poor and immature readers are insensitive to structure; good, immature readers do obtain meaning from text, and it seems unlikely that they are unaware of structural constraints.

Question III

Three models of the reading process and one of speech processing are

related to readability issues because they attempt to explain how readers utilize textual constraints to access meaning. The reading models are the Analysis by Synthesis (AS) model (Goodman, 1967, 1970; Hochberg, 1970; Smith, 1971), the Adaptive Process (AP) model (Gibson & Levin, 1975; Levin & Addis, 1979), and the Semantic Schemata (SS) model (Kitsch, 1972; Marshall & Glock, 1978; Meyer, 1975, 1977) whose name has been invented by this writer. The speech model was developed by Fodor, Bever, & Garrett (1974).

All three reading models claim that readers attend to a restricted set of available text cues because they possess internalized knowledge of the rule systems governing the distribution of text characteristics, and that the nature of the text search is influenced to varying degrees by the principle of selective attention and by the knowledge and experience of the reader. All the models posit that readers are sensitive to graphemic, syntactic, and semantic sentential constraints, but the SS model extends this description of perceptually salient text characteristics, emphasizing the importance of the semantic structure (logical and rhetorical relationships) of total discourse in text processing. Because research findings (cited above) provide evidence that readers utilize sentential and discourse constraints, it would seem that a synthesis of the models is desirable.

Both the AS and SS models claim that readers form hypotheses regarding the nature of the underlying structure of text (fully specified syntactic and semantic relationships) from attention to a reduced set of surface structure characteristics. That is, readers are able to determine meaning when they are attentive to a partial set of the semantic and syntactic attributes of a given unit of communication. This is roughly analogous to recognizing a particular person given only a silhouette. The AP model claims that readers do not use hypotheses in accessing meaning; instead, readers form what can be regarded

as a descriptive matrix of meaning as text is linearly processed. This matrix is built up as more information (text attributes) are added in linear processing. As the matrix of meaning becomes increasingly specified, the number of attributes that might be compatible with the extant features of the matrix decreases, and the reader searches the text for only compatible characteristics. Thus, economy of processing, not hypothesis-testing, is the principle that motivates the perceptual search.

The claims of the AS model would predict that the number of perceptual errors made by readers should decrease as sentences are linearly processed because the constraints imposed by the attributes of prior text decrease the number of attributes that could occur in subsequent text. Research findings provide counter-evidence to this prediction; there are more oral reading errors in the second fifth of clauses, and oral reading substitution errors are more often sensible through the point of substitution than beyond (Kolers, 1970), indicating that readers may form hypotheses early in the processing of the sentence which tend to be resistant to the effects of subsequent text cues; the eye-voice span, regarded as a measure of processing difficulty, is longest at the ends of sentences (Buswell, 1920; Fairbanks, 1937), suggesting that the integration of text cues to meaning and/or the accessing of meaning is delayed until the entire sentence is scanned rather than built up as the sentence is scanned as suggested in the AP model.

Hypothesis-testing as defined in the AS model is related to hypotheses made about meaning on the basis of surface structure characteristics. Somewhat different hypotheses, or heuristics, are suggested in the speech processing model of Fodor, Bever, and Garrett (1974). These are strategies that are used to assign surface structure phrases to deep structure sentoids. Fodor, Bever, and Garrett argue that the sentoid is the basic unit of processing. They suggest that readers employ a canonical sentoid strategy in which the first noun is perceived as the subject noun and the first verb is perceived as its predicate. The research evidence

(cited above) that major and minor phrase boundaries, sentoids, and clauses are perceptually real provide support for these claims. The utilization of a passive recognition strategy can also be inferred from research findings. The eye-voice span is longer after the verb and the word, by, in passives than after the verb and a preposition in actives with equivalent surface structure (Levin & Kaplan, 1970). There are more regressions in agent-deleted passives in which a prepositional phrase begins with by than in agent-deleted passives in which the prepositional phrase begins with a preposition other than by (Wanat, 1976). Thus, it may be that the canonical sentoid strategy is abandoned and replaced by a passive recognition strategy when the reader enccunters be + tense, the auxiliary form of passives, in text.

The different treatments of hypotheses in processing by the AS and SS models, and by the speech processing model seem to be complementary. The former fail to specify mechanisms other than sensitivity to the distributional properties of text (distinctive features and semantic structure) that suggest what the nature of hypotheses might be, and how hypotheses might operate to facilitate the accessing of meaning; the speech processing model attempts to do this.

It seems possible to develop a partial synthesis of the models if the principle that motivates the perceptual search suggested in the AP model is replaced by hypothesis-testing. The synthesized model would incorporate all the other features of the models, expanding the description of the salient text cues to include both sentential and discourse parameters, and including a tentative description of hypotheses readers may employ. The synthesized model brings into clearer focus questions regarding the relative importance of sentential and discourse constraints and their probable interaction. These questions have implications for readability because they suggest that it may not be possible to develop an optimal measure of readability prediction without indexing both

sentential and discourse constraints.

It is apparent that no model provides a description of the process through which a reduced set of surface structure cues are used to access fully specified meaning. Without such a description it seems difficult to form testable hypotheses regarding the effects of variation in cognitive factors (such as the ability to categorize) on the process that relates the output of the surface structure search to the generation or recognition of fully specified deep structure. Within-reader factors of this nature may play a significant role in readability prediction.

Relevance of the Findings from the Present Study to the Questions Addressed

This research was based on four assumptions inferred from the empirical findings and theoretical formulations discussed above. They were (1) readers do employ hypotheses (heuristics) in the perceptual text search, (2) hypotheses can be described with reference to the attributes of surface structure, (3) hypotheses could be inferred from examination of the nature of cloze errors associated with the discrimination of particular levels of the factors included in this study, and (4) it is possible to arrange hypotheses in a hierarchy of difficulty that is consonant with the complexity of the grammatical structures upon which the hypotheses operate (e. g. canonical structure).

Therefore, it was thought that the findings of this study would provide data that could be used to develop greater specification of the nature of the hypotheses employed by readers. Changes in the nature of hypotheses associated with variation in structural complexity (written level factor) might provide information that could be used to infer indices to improve the predictive accuracy of readability formulas. Changes in the nature of hypotheses associated with variation in the comprehension of passages (reading level factor) could provide indirect evidence that different formulas may or may not be needed to assess readability for good and poor readers. Changes associated with variation in knowledge of topic

would yield information that could indicate whether one readability formula might be used to predict the readability of texts written on different topics.

This study was an exploratory investigation of readability as an interactive process influenced by both text and reader characteristics. Its findings cannot provide final answers to any of the specific questions addressed because the issue of readability as an interaction is as complex as the reading process itself. It was hoped that by addressing the question posed by Gray and Leary 47 years ago, the findings from this study might generate questions that could be tested to develop better understanding of readability as a process and better means of measuring readability.

Methodology

Sample

The sample population consisted of the entire eighth grade of the Ogdensburg Free Academy, the only secondary school in Ogdensburg, N. Y. The final sample consisted of 239 students.

Materials

Topic choice was based on the results of pretest questionnaires, completed by students six weeks before the testing was done, on which students indicated the extent of their knowledge and of their interest in each of ten topics. Four sets of two parallel passages were written. In each set of parallel passages, a word in one passage was the same part of speech as the word in the same ordinal position in the other passage. The greatest difference in the average number of syllables per hundred words was 1.44 syllables, and the number of words and sentences in passage sets was identical as each written level. The topic of one passage in each set was that on which the largest number of students had indicated high knowledge and low interest (The Dangers of the Cigarette Habit), and the topic of the other was that on which the greatest number had indicated low knowledge and low interest (The Palestinian Liberation Organization (PLO)). It was decided

to use topics of low interest because interest has been shown to affect the extent of comprehension of text (Estes & Vaughn, 1973). One set of parallel passages was written at each of the fourth, sixth, eighth, and tenth written levels according to the Fry Readability Formula (1968).

The passages on each topic used in this study were constructed from the same core of facts. Although there are some facts in the upper written levels that are not in the lower, the primary differences across levels are in the difficulty of the vocabulary and in the structural complexity of the sentences.

Cloze tests were constructed on the passages, using every fifth word deletion. The first and last paragraphs of the passages were left intact to provide orienting structure. The number of deletions varied across passage levels from 51 to 56. Fifty-one deletions were scored to maintain an equal number of scored errors at each passage level.

Test Administration

Each student was given one set of parallel passages. A sample test passage was provided, and students were instructed to fill in all the blanks in the tests. Passage levels were randomly assigned, and topic presentation was counterbalanced. The tests were monitored by ten former teachers. An hour was provided to complete the tests.

Scoring

All scoring was done by the researcher. The passages were first scored to determine the total number of cloze errors, using exact scoring criteria. Then a sample of the protocols was scored to devise a scoring system which could be used to describe the effects of incorrect insertions on the structure and meaning of the text sentence. Finally, a third scoring was done using the system devised in the second, sample, scoring. Although Goodman's taxonomy could not be used in this research because much of it is not applicable to the ana ysis

of cloze errors, and because it cannot be used easily in research involving a large population (Goodman, 1969), his approach to error analysis provided guidelines for the construction of the scoring system used in this research. In the final scoring, each error was scored three times according to different criteria regarding the nature of the error: syntactic agreement with the word deleted (Phrase Structure Agreement Categories (PSAC)), meaning agreement with the word deleted (Meaning Agreement Categories (MAC)), and, where appropriate, the effect of the insertion on the structure and/or meaning of the sentence (Interpreted Error Categories (IEC)). There were nine categories in PSAC scoring and eight in MAC scoring. In both scorings, categories were created for indecipherable and multiword insertions. Effects scored in the PSAC analysis were insertions which did not alter the structure of the original sentence, which altered the structure within a minor phrase (this category consisted almost entirely of insertions of adjectives for articles and the inverse), and which altered the structure of the original sentence in any way other than in a minor phrase. Insertions for connectives and insertions of connectives were also given categories in the PSAC scoring. Effects scored in the MAC analysis were insertions whose meaning was nondistinct (nearly synonomous) from the deletions, whose meaning was distinct (different but sensible in context) in comparison with the deletions, and whose meaning was nonsensical in sentence context. PSAC and MAC scoring judgements were based on a reading of the entire and otherwise undeleted sentence. Obviously, the criteria used in the PSAC and MAC scorings failed to provide a description of the way in which meaning and structure were altered by insertions. The Interpreted Error Categories (IEC) were constructed for this purpose. Those errors which altered the structure of the sentence in any way except through the insertion of optional elements in minor phrases and/or which rendered the otherwise undeleted sentence nonsensical were given IEC scorings. Twenty Interpreted Error

Categories were created. Nineteen of these described some specific effect, and one category was provided for errors whose effects were uninterpretable or ambiguous. One category in this analysis, errors which violated selection restrictions, was further analyzed to attempt to determine the nature of the restrictions most frequently violated, but this scoring was not subjected to statistical analysis.

A student in the Speech Department at Temple University scored a sample of the protocols to determine scoring reliability. Pearson product-moment correlation was performed to determine the extent of scoring agreement. Scoring reliability for the PSAC and MAC analyses was fairly good, ranging from .56 to .99 for most categories. IEC scoring reliability was low, ranging from -.1 to .89. Poor reliability for the IEC scoring may have been a consequence of the fact that the means of many IEC categories were very low, and a difference of one in a category scoring could have had considerable impact on the size of the correlation.

Statistical Analyses

A 4 (written level) x 2 (topic) analysis of variance with repeated measures on the last factor was performed to determine whether comprehension was significantly affected by variation in structural complexity and by readers' knowledge of topic. Omega squared was performed to determine the relative contribution of these factors to comprehension. Another post hoc measure, the Tukey HSD test for comparison among the means was performed to determine which of the cell means were significantly different from the others.

One multivariate analysis of variance (Program CANOVA) was performed on each of the PSAC, MAC, and IEC analyses for each of the two topics to determine whether there was a significant effect on the number of cloze errors in the categories in the analyses as a result of variation in written level and reading level. Error analysis would have been enhanced had there been a computer program

available that could have been used to perform multivariate analysis with repeated measures on a research design with the number of levels of the factors and variables employed in this study. Shifts in error patterns across topics were inferred from comparisons of the results of the analyses performed on each of the topics. The analysis of variance performed on written level and topic did not result in an interaction, making this limitation less serious.

Discriminant analyses (SPSS) using the Wilks stepwise procedure were performed as post hoc analyses of the findings on the multivariate data. The purpose of these analyses was to determine the set(s) of variables (PSAC, MAC, and IEC categories) that discriminated best among the levels of the dependent variables, written level and reading level, or among levels formed by the intersection of written levels and reading levels. These analyses provided data used to infer changes in the nature of the errors made at different written and reading levels for each topic. The choice of the dependent variable in each discriminant analysis was dependent on the results of the multivariate analysis. When an interaction of factor effects occurred, one discriminant analysis was performed using groups formed by the intersection of written and reading levels as the dependent variable. This resulted in a sixteen cell matrix which had four reading levels across each of the four written levels. When no interaction occurred, two discriminant analyses were performed using groups formed by written levels as the dependent variable in one, and groups formed by reading levels as the dependent variable in the other.

Results

Effects of Written Level and Topic on Comprehension

Both written level and topic contributed significantly ($p < .001$) to comprehensibility. There was no interaction. Written level was found to account for 34.75% of the variance in comprehension, and knowledge of topic was found to account for .86% of the variance. The validity of texts of strength of association, used to compute the amount of variance accounted for by the factors, has been questioned when applied to the results of fixed-effects designs, such as that used

in this study (Glass & Hakstain, 1969). Tests of association also tend to undervalue the magnitude of the association when applied to a fixed-effects design with repeated measures (Dodd & Schultz, 1973). Therefore, all that can be concluded regarding the magnitude of the effects of the factors is that they are at least as great as those found in these analyses.

No significant differences were found between the cloze error means of passages at the fourth and the sixth written levels. All other comparisons among the means differed significantly ($p < .05$). The means and standard deviations are reported in Table 1.

Table 1

Cloze Error Means and Standard Deviations for Written Level and Topic

Topic	Written level									
	4		6		8		10		Levels Combined	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Dangers of the Cigarette Habit	30.88	5.75	30.85	5.80	35.98	5.30	40.33	3.53	34.49	6.49
The Palestinian Liberation Organization	31.10	5.71	31.23	4.76	38.04	4.79	41.30	4.98	35.35	6.74
Topics Combined	30.99	5.71	31.09	5.29	37.01	5.21	40.82	4.32		

Effects of Reading Level and Written Level on the Nature of Cloze Errors Associated with the Topic of Greatest Knowledge and the Topic of Least Knowledge

A significant difference among the means ($p < .001$) was found in every multivariate analysis of variance performed on the Phrase Structure Analysis Categories, the Meaning Agreement Categories, and the Interpreted Error Categories. For the topic of greatest knowledge, there was an interaction between the effects of written level and reading level for Phrase Structure Analysis Categories and Meaning Agreement Categories. There was no interaction of factor effects for the Interpreted Error Categories for the topic of least knowledge, but the significance level was neared ($p < .025$) for the interaction of effects for the topic of greatest knowledge. There were no significant interactions of effects in the PSAC and MAC

analyses performed on the topic of least knowledge.

Inferences regarding the magnitude of the changes in the category scores in the PSAC and MAC analyses could have been biased by the fact that every error was given a coding in these analyses; increases in the magnitude of category scores could have been the result of the fact that there was an increase in the total number of errors across the levels of the factors. Therefore, the means of all the categories in the PSAC and MAC analyses were expressed as percentages (called Mean Percentages or MPs) of the mean cloze score at each of the factor levels. It was not necessary to do this for the IEC analyses because only selected errors were given IEC codings.

Inspection of the results of the discriminant analyses performed on the three scorings indicated that specific types of errors were not associated with the discrimination of specific written levels, reading levels, or the groups formed by the intersection of these levels. Subsets of errors tended to discriminate among levels of the factors. Summary generalizations based on a synthesis of the results of the discriminant analyses are presented below. Findings from each of the analyses and the statistics on which they are based can be obtained from the researcher.

Summary Generalization (Across Reading Levels) - Decreasing reading level was, in general, associated with a narrowing of the text units whose well-formedness was either maintained or created; in good comprehension, errors tended to maintain or create sentoid integrity while altering elements within the sentoid; as comprehension decreased, errors tended to alter sentoid boundaries, creating new sentoid units, to create only well-formed phrases, and to violate the logical relations between sentoids and the selection restrictions within them.

Summary Generalization II (Across Written Levels) - (1) (a) When the topic was familiar and when comprehension of text was poor, increasing structural complexity was associated with an increase in the MPs of errors which maintain the structure of the undeleted text and create alternative readings of the sentence.

(b) When the topic was not familiar, increasing structural complexity was associated with an increase in the MPs of errors which alter structure and create nonsense, regardless of the level of comprehension. (2) (a) When the topic was familiar and when structure was altered and/or nonsense was created in the otherwise undeleted sentence, a slight increase in complexity (from level 4 to level 6) was associated with errors which maintain sentoid boundaries but alter phrases within sentoids; increasing complexity of text was associated with errors which alter sentoid boundaries, creating new sentoid units, which create only well-formed phrases, and which violate logical relations between sentoids and selection restrictions within them. (b) When topic was not familiar and when structure was altered and/or nonsense was created in the otherwise undeleted sentence, increasing complexity was associated with all of the types of errors which discriminated across levels in familiar text. There was no "dimension of easiness," a unique set of errors which discriminated relatively easy text from more difficult text. However, there was a "dimension of difficulty" which provided additional discrimination of written level six from written level eight that was not found for familiar text. This dimension is associated with errors which create only contiguous phrases which are meaningful.

Analysis of Selection Restriction Violations

Selection violations were scored according to the locus of the violation to determine what types of violation occurred most frequently. The initial scoring resulted in the creation of fifteen categories. A subsequent partial scoring indicated that the initial scoring was very unreliable; it is extremely difficult to determine what selection restrictions are violated when the error involves any combination of subject, verb, object relations because of such factors as passivization. Therefore, categories were collapsed to provide a relatively reliable but more general description of the nature of the violation of selection restrictions. Fifty-eight percent of these violations occurred among subject-verb-object relations, 25% between modifier-modified relations, and 17% between

the noun in a prepositional phrase and the noun in the noun phrase preceeding it.

Conclusions and Implications

Question I - Are there untested indices of difficulty that can be used to enhance the accuracy of readability formulas?

Conclusion 1 - New, untested indices of readability designed to measure the number and complexity of sentoids in passages can be constructed which may enhance the predictive accuracy of readability formulas. However, it is not clear to what extent the indexing of sentoid complexity without the indexing of the semantic relationships will improve readability prediction.

The findings provide additional evidence that readers employ a canonical sentoid strategy; sentoid integrity tended to be maintained or created at all reading levels and written levels. There is also some weak evidence to support the utilization of a passive recognition strategy. This is found in the fact that errors which could result from failure to recognize passivization provided very poor discrimination of levels.

Those errors whose effect was to create only a contiguous subject-verb unit, the most fundamental canonical unit, were among those which best discriminated across levels in only two analyses. Consideration of this and of the nature of the errors which discriminated best across factor levels (errors which resulted in the manipulation of phrase and/or sentoid boundaries but maintained or created sentoid integrity) suggest that the canonical sentoid strategy may be relatively complex; it may operate to assign not only fundamental canonical structure but also to assign adjective, adverb, and prepositional phrases to sentoids.

Errors which altered sentoid boundaries tended to truncate the sentoid in which the deletion occurred by attaching a phrase from the deletion sentoid to the previous sentoid. These errors might be attributed to the failure to recognize the verb of the embedded sentoid or to the failure to logically integrate the sentoids.

Consideration of these findings indicates that measures which index the number and complexity of sentoids may enhance readability prediction. Ratios, such

as the number of verbs divided by the number of nouns or of sentences, might provide such an index. However, it was not only errors involving the manipulation of phrase and sentoid boundaries that best discriminated among factor levels; errors which violated selection restrictions and clausal relationships were also among this set. Thus, it is not clear to what extent the indexing of sentoid complexity without indexing the complexity of logical relations within and between sentoids would improve readability prediction.

Among the difficulties inherent in attempting to index the complexity of selection restrictions and clausal relationships is the fact that interrelated structural and semantic relationships impact on these text factors (Chomsky, 1965; Meyer, 1977). It seems very unlikely that some ratio of an index of vocabulary load as currently measured in readability formulas (the frequency with which the words used in the texts being evaluated occur in text in general or the average number of syllables per word) to an index of structural complexity could capture the complexity of selection restrictions or clausal relationships.

Corollary 1 to Question I - Can a readability formula be constructed to be easily applied and more accurate than currently popular formulas?

Conclusion 2 - A more accurate readability formula would not be as easy to apply as currently popular formulas because it would involve the identification of at least particular phrase structure elements. Obviously, a formula requiring the counting of phrase structure elements would not be widely used unless the increase in predictive accuracy were great.

Corollary 2 to Question I - Can a readability formula be constructed to be applicable to all topics, regardless of the level of knowledge of the reader?

Conclusion 3 - If the magnitude of the effect of topic familiarity is no greater than that found in this study, the predictive accuracy of a readability formula should not be greatly diminished by the extent of a reader's knowledge of topic. If, however, the magnitude of the effect of topic familiarity were found, in subsequent research, to be greater, it would be expected that structural complexity would impact differently on the comprehension of familiar text than on the

comprehension of unfamiliar text.

When the topic was familiar and the text became increasingly complex, readers who were having difficulty made an increasing percentage of errors which maintained sentence structure and created a sensible alternative reading of the otherwise undeleted sentence, but when text was less familiar this did not occur. Thus, it appears that the impact of structural complexity on the processing of familiar text differs from its impact on the processing of less familiar text.

The comparisons discussed above occurred in the context of an increasing number of errors at each written level for both topics. Furthermore, the amount of the variance in comprehension scores accounted for by the topic factor was quite small. Therefore, the magnitude of the effect of topic familiarity on readability may be so small that it would be unnecessary to index it in a readability formula.

It is possible to interpret the findings discussed above in the context of a model developed by LaBerge and Samuels (1974) in which it is posited that readers switch their attentional focus from one kind of text cue to another as a function of the relative difficulty of the types of text cues. Readers who were having difficulty with the more complex text may have increased their attention to the structural aspects of the text to generate reasonable lexical insertions for the cloze blanks. They were able to do this for the topic of greatest knowledge because they possessed a background of knowledge that helped them make lexical insertions whose attributes were appropriate to the interrelated structural and semantic constraints (e. g. object noun that is + animate) imposed by extant text. They could not profit from these cues for the topic of least knowledge because they did not have sufficient background information to recognize the constraints imposed on the lexical insertions by extant text. That is, structure was a meaningful text cue for the topic of greatest knowledge because its interrelated semantic attributes were also meaningful. If this were the case, it provides further illustration of the interrelationship between the structural and semantic attributes of text.

Question II - Can the same formula be used to assess the readability of text for mature and immature readers and for good and poor readers?

Conclusion 4 - Although the findings from this study can be interpreted as evidence that different formulas are probably not needed to predict readability for good and poor readers, the evidence is indirect and inconclusive.

This study has the same failing as the Gray and Leary study; poor readers were identified as poor comprehenders (readers who did not comprehend the text well), not as poor readers in the traditional sense (readers whose reading achievement is below their general level of cognitive development). However, it is possible to generate a tentative conclusion concerning this question by consideration of the errors that discriminated across levels.

It was suggested in the discussion of Question I that poor comprehension is associated with the use of an heuristic which assigns phrases to sentoids without regard for the constraints imposed on the assignment of these phrases by verbs subsequent to the phrases, and with the failure to recognize or process selection constraints and clausal relationships. Assuming that these errors reveal the perceptual bias of poor readers as well as of poor comprehenders, it is possible to infer that the poor readers who participated in this study would be able to understand text which had fewer embeddings and a less difficult lexicon. Therefore, if the distribution of poor readers among the students who received the levels four and six passages was the same as the distribution of poor readers who received the levels eight and ten passages, the number of students whose scores fell at the independent and instructional reading levels should have been greater at written levels four and six than at the other levels. Of course, this was the case. This evidence is inconclusive because poor readers were not a manipulated factor in this study.

Question III - Does some one model of the reading process or some one synthesis of models present a better account of the findings on readability and perception?

Conclusion 5 - A synthesized model presents a better account of past findings and of those from this study. It should include aspects of both phrase structure and meaning relationships in a description of strategies (hypotheses or heuristics)

readers employ in accessing meaning, an account of the influence of within-reader factors on the nature of the strategies employed, and a description of the process through which the use of strategies facilitates the accessing of meaning.

Much of the argument for this conclusion was made in the background section. The results of this study add support to some aspects of the conclusion.

The finding that the sentoid appears to be the basic unit of processing and that the canonical sentoid strategy is complex, operating to assign phrases to sentoids, supports the use of hypothesis-testing in processing. The power of violation of selection restrictions and violation of clausal relationships to discriminate among factor levels indicates that both structural and semantic constraints should be included in a description of the cues to which readers attend. Additional evidence for the latter conclusion is found in the fact that knowledge of topic influenced the readers' ability to utilize text constraints in processing. Consideration of the findings of this study relative to the assumptions made before the study was done suggest that greater insight into the nature of the hypotheses readers may utilize might be gained through changes in miscue (error) analysis descriptions.

The assumption that the findings of this study could be used to infer a hierarchy of heuristics employed by readers was naive. It was expected that particular subsets of errors would discriminate specific factor levels from others; for example, it was expected that errors that created only fundamental canonical units would provide unique discrimination of the most complex written level. It is possible that these findings did not occur because the frequency and nature of the errors that could occur were constrained by the use of the cloze test. However, it seems more likely that hypotheses are considerably more complex than initially envisioned, and that failure to identify a hierarchy of hypotheses was related to the naivete of another assumption - that hypotheses can be described with reference to only the attributes of surface structure.

The results of this study and the findings of Kolers (1970) provide

convincing evidence that the processing of sentoids that occur at the beginning of sentences results in the perception of a sentoid unit (although the unit perceived and the text unit may not be the same) more frequently than the processing of sentoids that occur later in the sentence. One reason for this may be the fact that the surface structure of initially occurring sentoids is more frequently like their deep structure than the surface structure of subsequently occurring sentoids is like their deep structure. This explanation fails to take into account the fact that the deep structure of sentences with embeddings is more complex than the deep structure of simple sentences. Therefore, it seems that the relative complexity of the deep structure of sentences is more likely to be a causal factor in the findings than the relationship between the surface structure and the deep structure of discrete sentoids in a sentence, although both may be involved. In any case, error descriptions like those used in this study cannot capture the relationship between surface and deep structure characteristics. The following error sentence is taken from the protocols (the word deleted is underlined and the insertion is in parentheses).

> The land had been occupied but not owned (long) by
> Jewish and Arab peoples.

This error was coded in the Interpreted Error Category described as: Insertions for a verb which create a noun phrase, adverbial phrase, or prepositional phrase when joined to the last phrase in the previous sentoid, and render the deletion sentoid null or ungrammatical. Although this does describe general surface structure effects, it fails to describe the aspects of surface structure that may have been associated with the deep structure string that was generated by the error. A category description such as the following would have been more informative: Insertions whose effect was to create a single deep structure sentoid through the insertion of an adverb for the second verb in a passive sentence whose surface structure contained a compound verb phrase.

Similar effects on deep structure are found in the following error sentence taken from the protocols:

> Hostages are taken in the hijackings and they (wars) are killed unless ransom demands are met.

This error was coded in the Interpreted Error Category: Insertions which violate selection restrictions. It could have been better described as: Insertions whose effect is to create a single deep structure sentoid with a compound noun phrase in a prepositional phrase in the verb phrase through the insertion of a noun for the subject noun of the second sentoid in a compound sentence containing two passive sentoids. The alternate descriptions are superior to those used in this research for two reasons: first, the actual descriptions placed the errors in two categories although the errors are similar when their effects on deep structure are considered; second, the actual descriptions failed to indicate specific similarities in the surface structure of the two sentences (the presence of a connective and of passive markers) which may be perceptually salient characteristics triggering the assignment in the deep structure of sentences of this type of phrases from one sentoid to the verb phrase of the previous sentoid.

It is possible to infer from the findings of this research only the following hypothesis.

> Assign phrases subsequent to a verb to that verb phrase unless there is a second verb.

This hypothesis is minimally explanatory. Had the alternate error descriptions suggested above been found to discriminate between or among factor levels, utilization of the following, more explanatory, hypothesis could have been tentatively inferred.

> Assign phrases following connectives in surface structure verb phrases to the deep structure of that sentoid in the presence of passive markers.

One important implication of this research is that it does seem desirable to shift the focus of miscue analysis from the effects of errors on surface structure and general comprehensibility to a description of the deep structure strings associated with specific types of surface structure alteration. Research employing this type of analysis should help facilitate an understanding of the

process through which readers who attend to a reduced set of surface structure cues access fully specified meaning. This conclusion is based on the claim that fully specified deep structure is accessed through the use of strategies which operated on surface structure. This claim may turn out to be invalid; it is possible that decisions made about the nature of surface structure at the time that processing strategies are involved may generate supralinguistic descriptions such that the notion that fully specified deep structure strings are accessed in processing may have to be abandoned. The rejection of the claim that derivational complexity was a primary factor in text processing was partly responsible for the development of the claim that heuristics are employed in the recovery of deep structure (Fodor, Bever, and Garrett, 1974). If the latter claim is rejected after testing, it should lead to the generation of claims that will come closer to the realities of processing.

Other questions for future research were suggested earlier: (1) Will readability prediction be substantially improved through the inclusion of ratios which reflect the syntactic complexity of sentences? (2) Is the impact of reader knowledge of topic on text comprehension great enough to warrant its indexing in readability formulas?

The most important practical implication of this study is that because readability should be regarded as a dynamic process rather than a static attribute of text, people who use readability formulas should evaluate their results with reference to the characteristics of the population for which the readability analysis is done. Although it is not possible at this point to measure degrees of within-reader characteristics, such as knowledge of topic, or to provide a specific weight for such characteristics, common-sense evaluation of these characteristics may result in a more realistic appraisal of the readability of text for specific readers.

References

Abrams, K., & Bever, T. M. Syntactic structure modifies attention during speech

perception and recognition. Quarterly Journal of Experimental Psychology, 1969, 21, 280-290.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. Frameworks for comprehending discourse, Technical Report No. 12, Urbana, Illinois: University of Illinois, July, 1976.

Anderson, I. H., & Swanson, D. E. Common factors in eye movements in silent and oral reading. In C. A. Ruchmich (Ed.), Studies in the psychology of reading. Psychological Monographs, 1937, 48, 61-69.

Bernstein, M. R. Relationship between interest and reading comprehension. Journal of Educational Research, 1955, 49, 281-288.

Bever, T. G., Lackner, J. R., & Kirk, R. The underlying structures of sentences are the primary units of immediate speech processing. Perception & Psychophysics, 1969, 5, 225-231.

Blumenthal, A. L. Observations with self-embedded sentences. Psychonomic Science, 1966, 6, 453-454.

Bormuth, J. Readability: A new approach. Reading Research Quarterly, Spring, 1966, 79-132.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Buswell, G. T. An experimental study of the eye-voice span in reading. Supplementary Educational Monographs, 1920, 17.

Chall, J. S. The influence of previous knowledge on reading ability. Educational Research Bulletin, 1947, 26 (9), 225-230.

Chomsky, N. Aspects of the theory of syntax, Cambridge, Massachusetts: M.I.T. Press, 1965.

Coleman, E. B. Learning of prose written in four grammatical transformations. Journal of Applied Linguistics, 1965, 49, 332-341.

Dodd, D. H., & Schultz, R. F. Computational procedures for estimating the magnitude of effect for some analysis of variance disigns. Psychological Bulletin, 1973, 79, 391-395.

Estes, T. H., & Vaughn, J. L. Reading interest and comprehension implications. The Reading Teacher, 1973, 27, 194-253.

Fairbanks, G. The relation between eye-movements and voice in oral reading of good and poor readers. Psychological Monographs, 1937, 48, 78-107.

Fodor, J. A., Garrett, M. F., & Bever, T. G. Some syntactic determinants of sentential complexity, II: Verb structure. Perception and Psychophysics, 1968, 3, 453-461.

Fodor, J. A., Bever, T. G., & Garrett, M. F. The psychology of language. New York: McGraw-Hill, 1974.

Foster, K. I. Visual perception of repidly presented word sequences of varying complexity. Perception & Psychophysics, 1970, 8, 215-221.

Fry, E. A readability formula that saves time. The Journal of Reading, 1968, 11, 513-516.

Garrett, M. F., Bever, T. G., & Fodor, J. A. The active use of grammar in speech perception. Perception and Psychophysics, 1966, 1, 30-32.

Gibson, E. J., & Levin, H. The psychology of reading. Cambridge, Massachusetts: M.I.T. Press, 1975.

Glass, G. V., & Hakstain, R. A. Measures of association in comparative experiments: Their development and interpretation. American Educational Research Journal, 1969, 6, 403-413.

Goodman, K. S. Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 1967, 4, 126-135.

Goodman, K. S. Analysis of oral reading miscues. Reading Research Quarterly, 1969, 5, 9-30.

Goodman, K. Behind the eye: What happens in reading. In K. Goodman & O. S. Niles (Eds.), Reading Process and Program. Urbana, Illinois: National Council of Teachers of English, 1970.

Gray, W., & Leary, B. E. What makes a book readable. Chicago: The University of Chicago Press, 1935.

Hochberg, J. Components of literacy; Speculation and explanatory research. In H. Levin & J. P. Williams (Eds.), Basic studies in reading. New York: Basic Books, 1970.

Johnson, N. F. The psychological reality of phrase-structure rules. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 469-475.

Johnson, M. S., & Kress, R. A. Informal reading inventories. Newark, Delaware: International Reading Association, 1965.

Jones, M. B., & Pikulski, E. C. Cloze for the classroom. Journal of Reading.

Judd, C. H. Reading: Its nature and development. Chicago, Illinois: University of Chicago Press, 1918.

Kitsch, W. Notes on the structure of semantic memory. In E. Tulring & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Klare, G. R. The role of word frequency in readability. Elementary English, XLV, January, 1968, 12-22.

Kolers, P. A. Three stages of reading. In H. Levin & J. P. Williams (Eds.), Basic studies in reading. New York: Basic Books, 1970.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Levin, H., & Addis, A. B. The eye-voice span. Cambridge, Massachusetts: M.I.T. Press, 1979.

Levin, H., & Kaplan, E. L. Grammatical structure and reading. In H. Levin & J. P. Williams (Eds.), Basic studies in reading, New York: Basic Books, 1970.

Marshall, N., & Glock, M. D. Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. Reading Research Quarterly, 1978, 4, 10-56.

Meyer, B. J. F. Identification of the structure of prose and its implication for the study of reading and memory. Journal of Reading Behaviors, 1975, 7, 7-47.

Meyer, B. J. F. Organization of prose memory: Research with application to reading comprehension. In D. P. Pearson & J. Hansen (Eds.), Reading, Theory, Research, and Practice. Twenty-sixth yearbook of the National Reading Conference, Clemson, South Carolina: The National Reading Conference, Inc. 1977.

Robinson, F. P. The effect of language style on reading performance. Journal of Educational Psychology, 1947, 38, 149-156.

Schlesinger, I. M. Sentence structure and the reading process. The Hague, Netherlands: Mouton, 1968.

Smith, F. Understanding reading: A psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart, and Winston, 1971.

Tinker, M. A. Bases for effective reading. Mineapolis, Minnesota: University of Minnesota Press, 1965.

Vogel, M., & Washburne, C. W. An objective method of determining grade placement of children reading material. Elementary School Journal, January, 1928, 373-381.

Wanat, S. F. Relations between language and visual processing. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Washburne, C. W., & Morphett, M. V. Grade placement of children's books. Journal of Experimental Education, January, 1938, 355-64.

Weber, R. M. First-grades use of grammatical context in reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970, 147-163.